



Automatic transcription : Input and Output files formats, their compatibility and uses

To begin:

« What file formats can I send in transcription? » « How do I record my files and with what equipment? » , « Can a video be transcribed? » « Once I have transcribed, what are the formats that I can obtain? » « What are the differences and uses of each of these formats? ».

To answer all these questions, we decided to dedicate a folder on input and output files related to automatic transcription technology. It is also an opportunity to learn more about the Authôt application and all the services that we offer.

Have a good reading with Authôt !

Marylène REVER, Digital Marketing Manager
mrever@authot.com +33 (0)1 58 46 86 88

Table of contents

I- Automatic transcription: Know which recording formats are compatible	4
Audio recording formats for transcription.....	5
Video recording formats for transcription.....	6
Recording quality, essential for a good transcription!	7
II- Automatic transcription: Discover all the files formats downloadable and their uses	9
Different files formats for different uses	9
1- Text files formats .txt and .docx.....	9
2- Subtitle files formats .srt .webvtt.....	10
3- Files formats with timecode	11
4- html files formats	11
5- xml files formats.....	12
6- sjson files formats	12
Automatic transcription with Authôt application.....	12
III- Appendix.....	I3

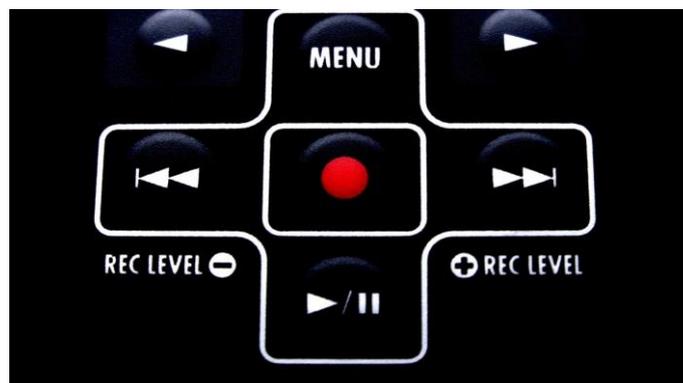
I-Automatic transcription: Know which recording formats are compatible



Before beginning our article, it seems essential to explain in a synthetical way what automatic transcription is. An automatic speech-to-text system is basically a collection of computer codes designed to convert a speech recording into its word-for-word text version. If you are interested in this topic, we invite you to read [this article](#) from our blog in the R&D section.

For saving time, a lot of Companies, Institutions, Medias, Universities and School use an automatic transcription system. So, which are the files compatible for a transcription? Just audio files?

Audio recording formats for transcription



Transcription is a simple process: Convert Speech to Text. When we say 'transcription', that's 'audio file' which comes to mind first. As we will see later on, recording quality is essential in order to obtain a reliable transcription.

Audio file formats are very numerous, some of them are a proprietary technology such as the .wma file or Waveform Audio File Format, developed by Microsoft.

Others are free and open as the .wav file or Waveform Audio File Format is a Microsoft and IBM audio file format standard for storing an audio bitstream on PCs. This format is compatible with Windows, Macintosh, and Linux operating systems. WAV files can also be edited and manipulated with relative ease using software.

If you have these extensions in your audio recordings library, or if you have unusual formats as .3gpa, don't panic! They will be accepted by the transcription system, provided that the recording quality is good, as we'll see later.

For example, on the online platform Authôt, you just have to send your audio format and the system will convert it automatically in .MP3 or .OGG. The conversion in .OGG is needed in order to be compatible with browsers which don't accept the mp3. For more informations, we invite you to consult the list of formats which are accepted on our platform, these are all present in the library FFMPEG (more details in the appendix).

Video recording formats for transcription



Audio files are not the only ones to be concerned by the transcription! Video files are too, and in this way it's the video's sound which is transcribed.

A numeric video, is a file which contains images, sound and text (metadata) placed in a container. In this container, images, sound and text are compressed. Compression and decompression of these files are realised by codec. A container is a file format which contain the audio and video stream, the codec information and the metadata. A codec (acronym for coding/decoding) is a compression / decompression algorithm of an numeric audiovisual signal.

Such as audio files, there is a multitude of different video files formats. For example, the .MOV, file extension compatible with QuickTime or, the .VQF which, in the line of the .MP3, allows a compression more important with a better quality.

Similar to audio files, every format are accepted by the transcription system. They will be automatically converted in .MP4, compatible with the several browsers.

For general information, on our transcription application, on average each month since the beginning of the year, our users send:

- audio files: 36%
- video files: 64 %

Video files are then, more and more important due to the development of a broad range of new services: captioning, synchronization, buning, translation etc.

Video is unquestionably the great winner on the web over the past few years!

Recording quality, essential for a good transcription!

The result of the transcription depend mainly on recording quality (our system has 95% accuracy with high quality audio/ video records).

Today, the majority of recorders are numeric¹. Their advantages: a compatibility with computers via an USB connection, a large memory capacity and a long autonomy.

All microphones do not capture sounds in the same manner, some microphones are designed to capture sounds coming from one single direction, and others are sensitive in all directions of space. As a consequence, the choice of the material is essential in order to have a good recording rendering, in particular if you are recording in a slightly noisy environment.

In general, you are highly advised to record in a quiet place without surrounding noises, indeed, microphones do not make differences between these noises and voice.



Thus, they will highly reduce the transcription quality. For example, a recording made in a restaurant will not have many chances to have a good quality because of the numerous external noises.

However, recording in a car with closed windows, regular speed, and without radio is a possible solution. Indeed, the overlap between car noise frequencies and voice frequencies are low and do not prevent from a good transcription.

To improve the quality of your recording, instead of speaking loudly, try to get closer from the microphone. Be careful, however, being closer from the microphone also increases the risk of saturating the output signal.

Elocution is also a major element concerning the recording, it should not be too fast. Tone of voice must be regular. Strong accents also impact on transcription quality.



Recordings can be made during meeting, in this case, it is recommended to place at least two microphones at the end of the table and connected with a recorder in order to capture everybody's voices.

Each person can also wear a lapel microphone, connected with a recorder. Other solution, the room can be equipped of a conference octopus. During meetings, you must avoid speakers interrupt themselves or speak at once. This strongly alters the transcription.

¹ <http://www.authot.com/en/nagra-digital-recorders/>

Thus, for a reliable transcription, a good recording is necessary! Authôt technology offers a multi-speakers and automatic transcription system, ideal for one-to-one conversation and meeting.

Save time sending your audio or video recording files at the format of your choice on app.authôt.com and obtain your transcription in one click!

An overview has been done on the main input files that can be sent to the Authôt application, for a complete list, we invite you to read our appendix at the end of this document.

The second part of our document will give you in detail the different formats available for export on the application once the transcription is done and their different utilities.

II- Automatic transcription: Discover all the files formats downloadable and their uses



Our previous article explained the automatic transcription goals and the different input formats you can send for transcription. Indeed, audio and video files are accepted for automatic transcription, as long as the recording quality is good. They both meet different objectives.

As a reminder, every sectors can be concerned by an automatic transcription service. This is why, a good comprehension of files formats is essential. Automatic transcription is the Speech to Text conversion. Input files, or "Speech" files have been seen, we offer you now to review the different output files formats, or "Text" files and their uses.

Different files formats for different uses

1- Text files formats .txt and .docx

Text in european alphabet are composed of characters in finite numbers (letters, punctuation signs, etc.). A "text" file is then a file including only a sequence of characters. The file format .txt available on the application allows you to export the transcription obtained in plain text. It can be open with Notepad for example. This format helps you to get your transcribed text easily.

The online Authôt text editor will enable you to put in form your text (underline, bold, italics, align, do paragraphs etc.). Thus, it is possible for you to work your text directly online and export it in .docx. This file format "doc", unlike to the "text" allows you to keep the layout and can be open with Word. So this file format is a real time saving!

With these text formats you have your transcription in a raw version, or in a reworked version thanks to the online editor.

This is particularly relevant if you are a company or an Institution which needs meeting or conference transcriptions. Names of the different speakers can be added if necessary.

2- Subtitle files formats .srt .webvtt

First at all, you have to know that a subtitle file is composed of 2 types of informations:

- the text to display
- for every line of text, the start and end timing

The file format .srt can be open with the text editor [Sublime Text](#) but also with [Aegisub](#) in order to modify the synchronization between subtitles and video if necessary.

```

28 7
29 00:00:25,840 --> 00:00:28,866
30 I call on Authôt
31 for translation
32
33 8
34 00:00:28,906 --> 00:00:31,900
35 and subtitling
36 of video productions
37
38 9
39 00:00:31,990 --> 00:00:33,346
40 that we make
41 at Condé Nast.

```

.webvtt format is also for subtitling, here are the main differences with .srt format:

- There is space for optional header data between the first line and the first cue
- Timecode fractional values are separated by a full stop instead of a comma
- Timecode hours are optional
- Cue settings allow the customization of cue positioning on the video

The two main subtitles formats are available on Authôt application.

Subtitles files formats are now essential for companies which published frequently short videos on social networks such as Facebook (Watch our [interview with Condé Nast](#)). Indeed, with the autoplay, videos are watched automatically and subtitling allows a better visibility for the video.

Authôt also offers: synchronization, burned subtitles, and translation. Thus, subtitles can be done in every languages.

Subtitles files formats are also increasingly used by Education with MOOC's, and Public Institutions in order to meet digital accessibility standards (Read [our article](#)).

It is therefore important to respect several rules if you wish to have a reliable and accessible subtitled video. For example, in Teletext, which is used to display subtitles on some broadcast platforms, line length is limited to 37 fixed-width (monospaced) characters, since at least 3

of the 40 available bytes are used for control codes. Guidelines for both platforms are summarised in the table below²:

Platform	Max length	Notes
BROADCAST	37 characters, reduced if coloured text is used	Teletext constraint
ONLINE	68% of the width of a 16:9 video and 90% of the width of a 4:3 video	The number of characters that generate this width is determined by the font used, the given font size (see fonts) and the width of the characters in the particular piece of text (for example, 'lilly' takes up less width than 'mummy' even though both contain the same number of characters).

Transcription allows you to subtitle your video file, thanks to .srt and .webvtt formats.

3- Files formats with timecode

On Authôt application you can have your transcription in a "Timecode(.text)" format. That allows you to open and read the timecode of your file.

```
00:00:29 --> 00:00:33
Portuguese subtitling : key element for culture, sport and
international relations

00:00:34 --> 00:00:40
This is clearly a booming for the Portuguese speaking
countries, it makes Portuguese subtitling attractive on
diverse topics:
```

Timecode files are extremely valuable notably for audiovisual sectors which edited movies. The automatic transcription and the export in .text are real time savings ([interview here](#)). They have their scripts and video rushes.

4- html files formats

HTML format or Hypertext Markup Language, is the standard markup language for creating web pages and web applications.

This format is available on Authôt application and allows you to add audio and text synchronisation for your podcasts for example. We invite you to read our [dedicated article](#) which will give you some tips about the use of .html files formats.

```
<show><span start="0.60" end="0.81" class="word">Authôt </span><span
start="1.27" end="1.47" class="word">unveils </span><span start="1.51"
end="1.61" class="word">its </span><span start="1.62" end="2.06" class="word">
secrets </span><span start="2.09" end="2.25" class="word">in </span><span
```

² <http://bbc.github.io/subtitle-guidelines/>

5- xml files formats

In computing, Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

Opposite to the html language, the xml permits to create its own tags, they are entirely customizable. Thus, the xml language is less rigid. Download in .xml format allows you to sort out and format easily your data.

For example with Authôt, your data sort out can be done via the tags “words” which contains an attribute “start” and an attribute “end”. Thanks to the XML, you can create an customizable display, create subtitles etc.

6- sjson files formats

In computing, JavaScript Object Notation or JSON is an open-standard file format that uses human-readable text to transmit data objects consisting of attribute–value pairs and array data types. JSON is promoted as a low-overhead alternative to XML as both of these formats have widespread support for creation, reading, and decoding in the real-world situations where they are commonly used.

```

{"start": [1420, 2660, 3670, 3750, 4360, 9960], "end": [2550, 3670, 3750, 4360], "text": ["Authôt
invite you to discover its ", "last news: portuguese subtitling! ", "Just like spanish,
galician and ", "catalan, portuguese is part of ", "latines languages. Portuguese is
  
```

Sjson format appears like a table divided in three parts: “start”, “end” and “text”. Unlike html for example, where the “start” and “end” are in the text in order to allow the synchronisation. Sjson meets very specific requests, but is available on our application.

Automatic transcription with Authôt application

Export
×

Choose your Format

<input type="checkbox"/> Text (.txt)	<input type="checkbox"/> Timecode (.text)
<input type="checkbox"/> Document (.docx)	<input type="checkbox"/> Xml (.xml)
<input type="checkbox"/> Subtitle (.srt)	<input type="checkbox"/> Original xml (.xml)
<input type="checkbox"/> Subtitle (.webvtt)	<input type="checkbox"/> Tst (.tst)
	<input type="checkbox"/> Sjson (.sjson)
	<input type="checkbox"/> Html (.html)

Authôt is a French start-up specialised in automatic transcription. We offer an online application to transcribe audio or video files into text.

A long list of files formats is available on app.authôt.com, in order to meet every specific needs!

You want another format? Please contact us :)

To conclude, we invite you to read the transcript times for files sent in English or French on our application. Please note, however, that the times indicated are dependent on the activity on the platform, if it is important at the time you send your file, the times of transcription can increase slightly.

Duration in minutes of the sent file	Transcript time
1 minute	32 seconds
3 minutes	53 seconds
5 minutes	1 min 17 sec
15 minutes	3 min 25 sec
30 minutes	6 min 24 sec
60 minutes	12 min 45 sec

Thank you for your attention!

Marylène REVER, Digital Marketing Manager
mrever@authot.com +33 (0)1 58 46 86 88

III- Appendix

Possible audio/video formats to send to the [Authôt application](#):

3ga,	g722,	dts,	nuv,
3g2,	gsm,	dv,	oga,
3gp,	gxf,	dv1394,	ogg,
3gp2,	h261,	dvd,	ogm,
3gpp,	h263,	mpeg1video,	ogv,
4xm,	h264,	mpeg2video,	ogx,
aac,	iff,	mpegts,	oma,
ac3,	ipod,	mpegtsraw,	oss,
act,	ivf,	mpegvideo,	pva,
adts,	latm,	mpg,	qcp,
adx,	lmlm4,	mpv,	r3d,
aif,	m1a,	msv,	raw,
aifc,	m2a,	mtv,	rcv,
aiff,	m4a,	mulaw,	riff,
alaw,	m4b,	mvi,	rl2,
alsa,	m4p,	dxa,	rm,
amr,	m4r,	ea,	roq,
amv,	m4v,	eac3,	rpl,
ape,	matroska,	ffm,	rso,
applehttp,	mj2,	film,	rtp,
asf,	mka,	flac,	rtsp,
asx,	mks,	flv,	sap,
au,	mkv,	vc1,	shn,
avi,	mlp,	vmd,	siff,
avs,	mlv,	voc,	smk,
awb,	mmf,	vox,	snd,
bink,	mov,	vqf,	sox,
bmv,	mp1,	w64,	spdif,
bwf,	mp2,	wav,	stream,
c93,	mp3,	weba,	svcd,
caf,	mp4,	webm,	swf,
cdg,	mpa,	wma,	thp,
d10,	mpc,	wmv,	ts,
daud,	mpc8,	wtv,	tta,
dct,	mpeg,	wv,	nsv,
dfa,	mxf,	xa,	nut,
dirac,	mxg,	xmv,	xwma
	dpx,	dss,	dvf,

Output text files from the [Authôt application](#):

txt	xml
docx	html
srt	text (timecode)
webvtt	sjson